

METODOLOGIES COMPUTACIONALS PER A L'ESTUDI DE MACROMOLÈCULES

F. JAVIER LUQUE¹ I MODESTO OROZCO²

1. Departament de Fisicoquímica, Facultat de Farmàcia, Universitat de Barcelona.

Av. Diagonal, s/n. 08028 Barcelona.

2. Departament de Bioquímica i Biologia Molecular, Facultat de Química, Universitat de Barcelona.

Martí i Franquès, 1. 08028 Barcelona.

SUMMARY

Theoretical methods for the description of macromolecular systems are reviewed. Methods are classified according to the level of accuracy in the description of the atomic system. Quantum mechanical and classical methods are compared not only from a methodological point of view, but also from a «functional» point of view. The potential use of these techniques in the study of the structure and function of proteins is discussed.

INTRODUCCIÓ

L'objectiu de les tècniques computacionals aplicades a la biologia és proporcionar una comprensió dels processos que tenen lloc en els éssers vius d'acord amb les lleis i els principis de la química i de la física. La grandària i la complexitat dels sistemes biomoleculars, això no obstant, dificulten la consecució d'aquest objectiu. A més, cal tenir present que les macromolècules biològiques són sistemes dinàmics i, per tant, no és possible aprofundir en els fenòmens biològics sense considerar el paper exercit per les

fluctuacions estructurals. Tanmateix, cal considerar que els moviments macromoleculars varien en un marge molt ampli, des de centèsimes d'àngstrom fins a desenes d'àngstrom, i que l'escala temporal d'aquestes fluctuacions també és molt àmplia, des de femtosegons fins a segons. És obvi, doncs, que l'estudi dels processos biològics requereix una gamma diversa de mètodes computacionals.

En els darrers quaranta anys les tècniques computacionals han adquirit una posi-

ció prominent dintre de la ciència, a causa del vertiginós augment en la potència de càlcul dels ordinadors, que actualment s'incrementa un ordre de magnitud cada cinc anys. Tanmateix, aquesta evolució sembla seguir la mateixa progressió, de manera que els recursos computacionals disponibles en un futur seran sensiblement superiors en els anys vinents. Per tant, cal esperar que la importància de les tècniques de simulació per ordinador esdevingui més gran, no tan sols pel que fa al nombre de treballs publicats, sinó sobretot per la possibilitat d'analitzar sistemes biomacromoleculars cada cop més grans, i una gamma de processos biològics més àmplia.

El propòsit del present capítol és oferir una visió sobre les diverses tècniques computacionals disponibles actualment per a l'estudi dels processos que tenen lloc en el si de les biomolècules. Cal destacar que no pretenem fer una anàlisi detallada dels diversos mètodes de simulació, sinó que considerem més adient proporcionar una descripció global del ventall de metodologies, destacant llurs característiques principals i llurs aplicacions en el camp de les macromolècules.

DEFINICIÓ DEL MODEL MOLECULAR

La simulació per ordinador d'un sistema químic, bioquímic o farmacològic comporta la consideració de tres factors diferents: a) el procés o la propietat del sistema real que ens interessa examinar; b) la precisió requerida per avaluar aquesta propietat; i c) els recursos computacionals disponibles. L'acció conjunta d'aquests factors condiciona la definició del model molecular que representarà el sistema real (vegeu figura 1). Aquesta és una etapa clau, ja que la qualitat dels resultats obtinguts amb el model escollit determinarà la comprensió del procés o de les

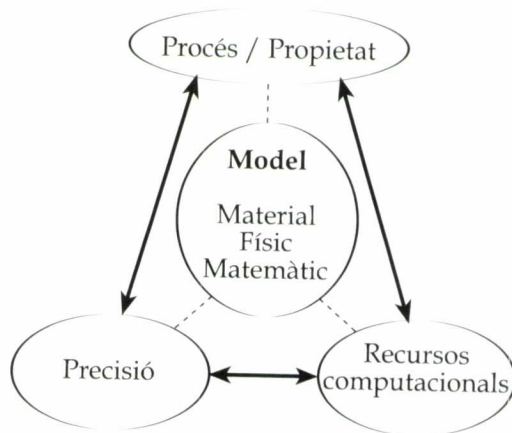


FIGURA 1. Factors que determinen la definició del model emprat en simulacions teòriques.

proprietats del sistema. La definició del model molecular es realitza considerant tres aspectes:

a) **Material.** Defineix la porció de matèria explícitament considerada en el model molecular. Atès l'elevat nombre d'àtoms en les biomolècules, i fins i tot en molts sistemes d'interès purament químic, cal considerar només aquella part que és essencial per descriure el procés. Per exemple, en el cas d'una reacció enzimàtica, el model material més simple estaria constituït per aquells residus directament que intervenen en el procés catalític. La selecció del model material defineix els graus de llibertat del sistema. Així, si el nostre objectiu és l'estudi d'una reacció química, el model molecular haurà de considerar explícitament els nuclis i els electrons; ara bé, si l'objectiu fos l'anàlisi de la flexibilitat conformational d'una proteïna, seria suficient fer-ne una descripció atòmica.

b) **Físic.** Un cop seleccionat el model material, hem de definir quines lleis regulen la interacció entre les partícules, és a dir, quina relació hi ha entre l'energia del sistema i els graus de llibertat de les partícules. En aquest punt, caldrà escollir, per exemple, entre l'e-

quació de Schrödinger o les equacions clàssiques del moviment formulades per Newton.

c) **Matemàtic.** Defineix les característiques tècniques i computacionals per al tractament del model material i físic. Dintre d'un estudi de mecànica quàntica, la selecció de la base i del nivell de teoria emprat per obtenir la funció d'ona formarien part del model matemàtic. Altres components d'aquest model serien, per exemple, la selecció d'algorismes d'optimització de geometria o bé d'exploració de l'espai conformational de proteïnes o d'àcids nucleics.

Un cop disposem dels resultats de la simulació, cal finalment analitzar-los amb una doble finalitat. D'una banda, cal proporcionar una descripció precisa del procés d'interès, que permeti relacionar les noves dades amb les propietats estructurals i químiques, i a més oferir una interpretació conduent a un aprofundiment en les bases moleculars del procés. D'altra banda, cal examinar la validesa de les hipòtesis formulades en el moment de seleccionar el model material, físic i matemàtic, per tal de detectar possibles errades o inconsistències degudes a l'elecció del model. En aquest darrer cas, cal formular unes noves hipòtesis de treball i corregir el model molecular.

La comparació entre les propietats mesurades experimentalment i les dades teòriques és la prova final que determina la fiabilitat de la simulació computacional. No obstant això, cal tenir present que una bona concordança no implica necessàriament la validesa del model teòric. Així, és possible que tingui lloc una compensació d'errors, o bé que la propietat d'interès sigui relativament insensible a la selecció del model molecular. És precisament la comparació ponderada entre experiment i teoria la que afavoreix una progressió més detallada en la naturalesa dels processos biològics. Cal remarcar en aquest punt que la finalitat dels càlculs de simulació computacional no és

substituir la medicció experimental, sinó proporcionar uns mitjans *complementaris* a l'experiment, ja que l'obtenció de determinades dades per experimentació pot ser impossible o bé extremadament costosa. Aquesta complementarietat és, doncs, la veritable potència que aporten les tècniques computacionals.

CLASSIFICACIÓ DE LES TÈCNiques COMPUTACIONALS

Tal com es va esmentar anteriorment, la diversitat de processos i propietats d'interès per a la comprensió dels éssers vius requereix una gamma variada de tècniques computacionals. No obstant això, aquestes eines poden classificar-se en tres grans grups d'acord amb el model físic escollit: a) tècniques quàntiques; b) tècniques clàssiques; i c) tècniques mixtes.

Les tècniques quàntiques consideren explícitament els graus de llibertat nuclear i electrònics. En canvi, les tècniques clàssiques no consideren els electrons, ja que els àtoms que componen el model material són descrits com un conjunt de partícules que interaccionen entre si d'acord amb les lleis clàssiques. En el primer cas és necessari recórrer a lleis de la mecànica quàntica per descriure el sistema químic, mentre que en el segon cas el moviment de les diverses partícules és descrit mitjançant la mecànica clàssica. Des d'un punt de vista computacional, el nombre d'àtoms (nucli i electrons) que poden compondre el model molecular quan es fan servir les tècniques quàntiques és reduït, unes poques desenes d'àtoms, mentre que en el segon cas es pot tractar de models formats per milers de partícules. Malgrat aquesta limitació quant a la grandària del model, les tècniques quàntiques proporcionen una descripció del procés químic o de les propietats molecu-

lars més acurada que les tècniques clàssiques.

Les tècniques mixtes es van desenvolupar amb la intenció de combinar els avantatges propis de les tècniques quàntiques de les clàssiques. Així, doncs, la part del sistema que té més rellevància per a l'estudi del procés químic o biològic es descriu des del punt de vista mecanoquàntic fent servir un model material reduït, mentre que es fa servir un tractament clàssic per a la resta del sistema. D'aquesta manera, s'intenta optimitzar la precisió en la descripció de processos reactius o en l'avaluació de les propietats del sistema, alhora que la representació clàssica de l'entorn permet augmentar la grandària del sistema i considerar la seva influència sobre el procés o les propietats d'interès.

Seguidament es farà una revisió de les característiques principals d'aquests tres grups de tècniques computacionals, destacant les seves aplicacions.

TÈCNIQUES QUÀNTIQUES

Mètodes *ab initio*

Dintre del marc de la mecànica quàntica, la descripció d'un sistema molecular ve donada per la seva funció d'ona, que conté tota la informació assequible sobre el sistema. A fi de conèixer el valor d'una determinada propietat (*observable*), només cal aplicar el corresponent *operador* a la funció d'ona. L'objectiu essencial d'aquestes tècniques, doncs, és determinar, per a un conjunt de

nuclis i electrons, quina és l'expressió de la funció d'ona, $\Psi = \Psi(\mathbf{R}, \mathbf{r}, t)$, que depèn de les coordenades nuclears (\mathbf{R}) i electròniques (\mathbf{r}), i del temps (t).

L'obtenció de la funció d'ona comporta la resolució de l'equació de Schrödinger dependent del temps (equació 1). Aquesta és una tasca formidable, i afortunadament es pot simplificar moltíssim perquè molts sistemes d'interès químic són conservatius, cosa que permet obviar la dependència temporal. Aleshores, la funció d'ona pot obtenir-se resolent l'equació de Schrödinger independent del temps (equació 2), on H és l'operador hamiltonià i E és l'energia total del sistema.

$$\hat{H}\Psi(\mathbf{R}, \mathbf{r}, t) = -i \frac{\hbar}{2\pi} \frac{\partial \Psi(\mathbf{R}, \mathbf{r}, t)}{\partial t} \quad [1]$$

$$\hat{H}\Psi(\mathbf{R}, \mathbf{r}) = E\Psi(\mathbf{R}, \mathbf{r}) \quad [2]$$

L'operador hamiltonià (eq. 3; en unitats atòmiques) comprèn la suma de les energies cinètiques de nuclis i electrons, l'energia electrostàtica d'atracció nucli-electró, i les energies electrostàtiques de repulsió nucli-nucli i electró-electró. En l'equació 3, els índexs A, B fan referència als nuclis, mentre que i, j descriuen els electrons, i M i N són el nombre total de nuclis i d'electrons. A més, l'operador de segon ordre laplaciana ∇^2 representa la derivada respecte a les coordenades, m_A és la relació de la massa del nucli A amb la massa de l'electró, i Z_A és el nombre atòmic del nucli A . Finalment, \mathbf{R}_A i \mathbf{r}_i representen els vectors posició del nucli A i de l'electró i .

$$\hat{H} = -\sum_{A=1}^M \frac{1}{2m_A} \nabla_A^2 - \sum_{i=1}^N \frac{1}{2} \nabla_i^2 - \sum_{A=1}^M \sum_{i=1}^N \frac{Z_A}{|\mathbf{R}_A - \mathbf{r}_i|} + \sum_{A=1}^M \sum_{A>B}^M \frac{Z_A Z_B}{|\mathbf{R}_A - \mathbf{R}_B|} + \sum_{i=1}^N \sum_{j>1}^N \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \quad [3]$$

Obtenir la funció d'ona a partir de l'equació 2 és encara molt complicat. Per això, s'apliquen aproximacions com la de Born-Oppenheimer, que consisteix a deslligar el moviment nuclear de l'electrònic, ja que la diferència de masses entre nucli i electró permet considerar que els electrons poden adaptar la seva distribució en cada moment a la configuració nuclear. Aquesta aproximació permet definir l'hamiltonià electrònic, H_{el} , i obtenir la funció d'ona dels electrons, $\Psi_{el} = \Psi_{el}(\{\mathbf{R}\}, \mathbf{r})$, per a una determinada configuració nuclear $\{\mathbf{R}\}$ (equació 4). En aquest cas, l'hamiltonià electrònic únicament conté els termes corresponents a l'energia cinètica dels electrons, l'atracció electrostàtica nucli-electró i la repulsió coulòmbica entre electrons, encara que s'hi pot afegir el terme de repulsió entre nuclis, ja que és constant per a cada disposició dels nuclis (equació 5). Cal remarcar que tant la funció d'ona com l'energia electrònica depenen paramètricament de les coordenades nuclears.

$$\hat{H}_{el} \Psi_{el}(\{\mathbf{R}\}, \mathbf{r}) = E_{el}(\{\mathbf{R}\}) \Psi_{el}(\{\mathbf{R}\}, \mathbf{r}) \quad [4]$$

$$\hat{H}_{el} = -\sum_{i=1}^N \frac{1}{2} \nabla_i^2 - \sum_{A=1}^M \sum_{i=1}^N \frac{Z_A}{|\mathbf{R}_A - \mathbf{R}_i|} + \sum_{i=1}^N \sum_{j>1}^N \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \quad [5]$$

Un cop s'ha resolt l'equació 4, es pot abordar posteriorment el problema del moviment nuclear. La idea essencial és substituir les coordenades explícites dels electrons en l'equació 3 pels seus valors mitjans, mesurats sobre la funció d'ona electrònica. Això permet construir un hamiltonià nuclear efectiu per al moviment dels nuclis en el camp mitjà dels electrons.

El problema fonamental és, doncs, obtenir la funció d'ona dels electrons en moviment dins del camp elèctric creat pels nuclis fixos. El primer pas per assolir aquest objec-

tiu és l'aproximació de Hartree-Fock, que permet determinar els millors espins orbitals que minimitzen l'energia del sistema. Aquest tractament matemàtic condueix a una equació de valors propis (equació 6), on $f(i)$ és un operador monoelectrònic efectiu, l'operador de Fock (equació 7), $\{\chi\}$ el conjunt d'espins orbitals i $\{\varepsilon\}$ el conjunt d'energies orbitals. El determinant de Slater format per aquests espins orbitals és la funció d'ona Hartree-Fock i és la millor aproximació monodeterminant de l'estat fonamental del sistema. L'aproximació de Hartree-Fock, doncs, permet transformar el tractament d'un sistema de molts electrons en un problema monoelectrònic en què es considera la repulsió mitjana, $v^{HF}(i)$, entre els diferents electrons.

$$f(i) \chi(r_i) = \varepsilon \chi(r_i) \quad [6]$$

$$f(i) = -\frac{1}{2} \nabla_i^2 - \sum_{A=1}^M \frac{Z_A}{|\mathbf{R}_A - \mathbf{r}_i|} + v^{HF}(i) \quad [7]$$

Des del punt de vista pràctic, volem destacar dues característiques fonamentals del tractament Hartree-Fock. D'una banda,

els espins orbitals es descriuen com una combinació lineal d'orbitals atòmics (aproximació MO-LCAO). Aleshores, com més gran i complet sigui el conjunt d'orbitals atòmics, major serà la flexibilitat en l'expansió dels espins orbitals, i menor serà l'energia del sistema. En el cas de poder utilitzar una base d'orbitals atòmics infinita, es podria determinar el valor límit de l'energia Hartree-Fock. D'altra banda, ja que cada electró es mou dins el camp mitjà creat per la resta d'electrons, les correlacions instantànies entre electrons no es conside-

ren explícitament. Per introduir aquest efecte, és necessari prescindir de la descripció monodeterminant de la funció d'ona, que hauria de descriure's com una combinació dels determinants de Slater corresponents a l'estat fonamental i els estats excitats del sistema. Això permet definir l'energia de correlació electrònica, E_{corr} , com la diferència entre l'energia exacta no relativista, $E_{\text{no-rel}}$, i l'energia límit Hartree-Fock, $E_{\text{HF-lim}}$ (equació 8).

$$E_{\text{corr}} = E_{\text{no-rel}} - E_{\text{HF-lim}} \quad [8]$$

D'acord amb el que s'ha exposat prèviament, la qualitat dels resultats obtinguts per aquestes tècniques dependrà de la flexibilitat de la base d'orbitals atòmics i del nivell emprat per introduir la correlació electrònica (vegeu figura 2). Òbviament, com millor sigui la base d'orbitals i el tractament dels efectes de correlació electrònica, millors seran els resultats. No obstant això, aquest fet comporta un increment

enorme en l'esforç computacional, cosa que obliga a tractar models materials relativament reduïts, de manera que a la pràctica és necessari optimitzar el nivell de càlcul i la precisió desitjada en l'avaluació de les propietats moleculars amb els recursos computacionals.

A fi de reduir l'enorme cost computacional i d'augmentar la grandària del sistema, s'han proposat diverses estratègies i formalismes alternatius, alguns dels quals s'exposen a continuació.

Mètodes semiempírics

Aquests mètodes pretenen reduir l'elevat cost computacional dels mètodes *ab initio* mitjançant la definició d'un hamiltonià efectiu, introduint tota una sèrie de simplificacions en el tractament matemàtic dels mètodes *ab initio*, i suplint aquestes simplificacions amb la parametrització empírica de diversos termes, de manera que els resultats reproduïxien les dades experimentals per a un conjunt determinat de molècules. Les simplificacions més importants són les següents:

a) **Aproximació de «core».** Consisteix a considerar exclusivament els electrons de les capes de valència, suprimint el tractament explícit dels electrons situats en les capes internes, que conjuntament amb el nucli formen el *core* de l'àtom. Aleshores, aquests mètodes fan servir un hamiltonià electrònic efectiu de valència, on aquests electrons es mouen dins el camp creat pel conjunt de *cores* fixos.

b) **Base mínima.** La descripció dels electrons de valència es realitza fent servir una base mínima d'orbitals atòmics, és a dir, cada espín orbital molecular es representa amb una base formada pel nombre mínim d'orbitals atòmics necessaris per descriure els orbitals ocupats de l'àtom.

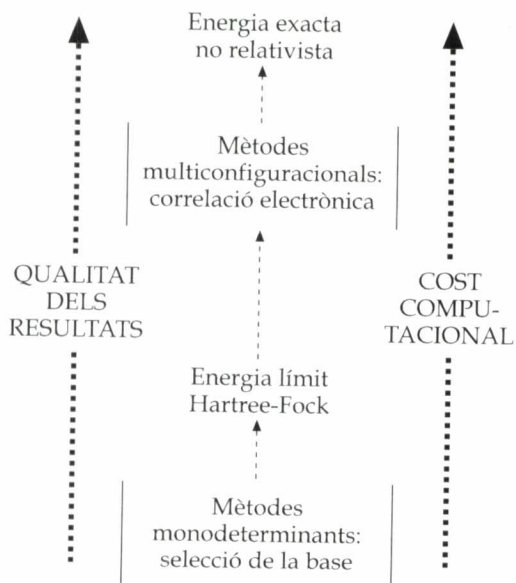


FIGURA 2. Representació esquemàtica dels nivells de teoria en els mètodes *ab initio*

c) **Aproximació de recobriment diferencial nul.** Consisteix a considerar que el recobriment diferencial de dos orbitals atòmics és zero si aquests dos orbitals són diferents (equació 9). Aquesta aproximació comporta una simplificació molt gran en l'avaluació del terme $v^{\text{HF}}(i)$ (vegeu equació 7). Cal indicar, no obstant això, que segons el mètode considerat, l'aplicació d'aquesta aproximació es realitza en un sentit més o menys estricte.

$$\phi_{\mu}^*(1) \phi_{\nu}(1) dr_1 = 0 \quad \text{si } \mu \neq \nu \quad [9]$$

Aquestes simplificacions, especialment la del recobriment diferencial nul, són prou dràstiques i, si bé augmenten sensiblement l'eficàcia computacional, comprometen seriosament la qualitat dels resultats obtinguts. Per evitar això, és necessari ajustar diversos paràmetres fent servir dades obtingudes amb mètodes *ab initio* o més generalment dades experimentals, com geometries d'equilibri, moments dipolars, calors de formació i potencials de ionització, entre d'altres. A més, hom preveu que el procés d'ajust d'aquests paràmetres permeti introduir implícitament els efectes de correlació electrònica.

Per raó de la parametrització, cal esperar que aquests mètodes reproduïxin les propietats moleculars no tan sols per a les molècules considerades en el procés d'ajust, sinó també per a les molècules estructuralment relacionades. No obstant això, no hi ha garantia sobre el comportament d'aquests mètodes per a molècules diferents. A més, en el

Mètodes de densitat funcional

Aquests mètodes es diferencien dels anteriors en el fet que la principal quantitat d'interès no és la funció d'ona, sinó la densitat electrònica. El principal avantatge computacional rau, doncs, en el fet que la descripció d'un sistema de N electrons, en lloc de fer-se amb una funció d'ona de dimensió $3N$, només necessita una funció, la densitat electrònica, que depèn de tres coordenades espacials.

Els mètodes de densitat funcional es basen en el teorema de Hohenberg-Kohn, segons el qual l'energia total d'un sistema en el seu estat fonamental és un funcional de la densitat electrònica del sistema, i que qualsevol altra densitat diferent de la densitat verdadera condueix a una energia més elevada. D'acord amb aquest teorema, el procediment per determinar la densitat consisteix simplement a minimitzar el funcional energia, $E[\rho(\mathbf{r})]$. Aquest funcional comprèn diversos termes (equació 10), que corresponen a l'energia cinètica dels electrons no interaccionants amb la mateixa densitat $\rho(\mathbf{r})$ que el sistema real d'electrons, $T[\rho(\mathbf{r})]$, l'energia electrostàtica entre els electrons amb un potencial extern (corresponent generalment a la distribució dels nuclis) i a la repulsió entre electrons, $U[\rho(\mathbf{r})]$ (equació 11), i finalment el terme $E_{\text{xc}}[\rho(\mathbf{r})]$, que conté les contribucions d'intercanvi i de correlació electrònica, a més d'incloure la diferència entre $T[\rho(\mathbf{r})]$ i la verdadera energia cinètica dels electrons del sistema.

$$E[\rho(\mathbf{r})] = T[\rho(\mathbf{r})] + U[\rho(\mathbf{r})] + E_{\text{xc}}[\rho(\mathbf{r})] \quad [10]$$

$$U[\rho(\mathbf{r})] = \sum_{A=1}^M \int -\frac{Z_A \rho(\mathbf{r})}{|\mathbf{R}_A - \mathbf{r}|} d\mathbf{r} + \frac{1}{2} \iint \frac{\rho(\mathbf{r}) \rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' \quad [11]$$

cas que s'obtinguin resultats erronis, no hi ha un procediment adient per millorar el comportament d'aquests mètodes.

El procediment pràctic per minimitzar l'energia i obtenir la densitat del sistema va ser proposat per Kohn i Sham i es basa en la

descripció de la densitat a partir d'un conjunt d'orbitals moleculars. Si bé no entrarem en més detalls, cal dir que, encara que el formalisme de Kohn i Sham és en principi exacte, la seva aplicació és inexacta ja que no es coneix la dependència funcional exacta del terme d'intercanvi-correlació amb la densitat. Actualment les diverses definicions disponibles permeten classificar els mètodes de densitat funcional en dues categories: a) local, quan aquest terme és funció única de la densitat; i b) no local, quan depèn, a més, del gradient de la funció densitat. Altres mètodes, tanmateix, incorporen el terme d'intercanvi determinat des del punt de vista Hartree-Fock en la descripció del funcional d'intercanvi electrònic.

És important tenir present la inexactitud del terme $\text{Exc}[\rho(\mathbf{r})]$ amb vista a l'aplicació d'aquests mètodes per al càlcul d'estructures i propietats de sistemes moleculars, ja que la seva precisió depèn directament de les aproximacions fetes en la construcció d'aquest funcional. En aquest sentit, doncs, la teoria de Kohn-Sham no és variacional respecte a l'energia de l'estat fonamental, encara que sí que ho és respecte al sistema model emprat per definir el terme $\text{Exc}[\rho(\mathbf{r})]$. Per comparació amb els mètodes *ab initio*, es podria dir que els mètodes basats en l'equació de Schrödinger proporcionen una aproximació a la funció d'ona real un cop conegut l'hamiltonià exacte del sistema, mentre que els mètodes de densitat funcional proporcionen la densitat òptima per a un funcional aproximat.

Aplicacions i limitacions

Magrat les diferències en el formalisme i el tractament matemàtic dels mètodes ressenyats anteriorment, tots ells descriuen un sistema químic considerant els graus de llibertat de nuclis i electrons. Per tant, aquestes són les eines computacionals que cal fer ser-

vir per analitzar qualsevol procés en què es produeixi un canvi o una reorganització electrònica, és a dir, la formació i el trencament d'enllaços covalents, i en general per a l'estudi de la reactivitat química. En aquest context, la informació proporcionada per aquests mètodes és molt diversa, si bé s'especifica de forma esquemàtica a continuació:

a) **Paràmetres geomètrics.** Es pot determinar la geometria de mínima energia caracteritzant les coordenades internes de la molècula. Més important és el fet que pot obtenir-se la geometria d'estats de transició, que no són assequibles amb tècniques experimentals, i es poden comprendre els mecanismes de reacció, com també es pot aprofundir en factors cinètics de les reaccions químiques i, en particular, es poden analitzar els factors que determinen l'eficiència catalítica dels enzims.

b) **Paràmetres energètics.** La resolució de l'equació de Schrödinger o de Kohn-Sham permet avaluar l'energia del sistema i construir les superfícies d'energia potencial dels processos reactius. Mitjançant el pertinent tractament per mecànica estadística dels termes corresponents, de les vibracions, de la rotació i de la translació moleculars, poden estimar-se diverses magnituds termodinàmiques, com l'energia lliure, que al seu torn permet avaluar l'equilibri termodinàmic entre reactius i productes; o bé es pot examinar la població de diferents conformers o tautòmers. Aquests estudis tenen rellevància per a l'estudi del procés d'unió entre molècules, i per a la identificació d'espècies moleculars bioactives.

c) **Propietats moleculars.** El coneixement de la funció d'ona permet avaluar qualsevol element observable a partir del seu corresponent operador, com per exemple el moment dipolar. Amb vista a l'estudi de sistemes bioquímics i farmacològics, cal remarcar l'aplicació d'algunes d'aquestes

propietats com a índexs de reactivitat, que permeten identificar els determinants moleculars que modulen el reconeixement entre substàncies, o bé la definició de grups farmacofòrics. Entre aquestes propietats, cal fer especial esment del potencial electrostàtic molecular, que correspon al valor esperat de l'operador \mathbf{r}^{-1} , i que proporciona l'energia d'interacció entre la distribució de càrrega molecular amb la unitat de càrrega positiva. La identificació de regions de potencial positiu i negatiu pot relacionar-se amb zones susceptibles a l'atac de nucleòfils i electròfils, respectivament. Tanmateix, els mínims de potencial electrostàtic poden relacionar-se amb la capacitat d'actuar com a acceptors de pont d'hidrogen, o bé amb les afinitats protòniques. Altres índexs derivats del potencial electrostàtic molecular, com el potencial d'interacció molecular generalitzat, proporcionen una descripció més precisa de la reactivitat molecular. Altres propietats emprades com a índexs de reactivitat són la densitat de càrrega, que pot distribuir-se entre els diversos àtoms a fi d'obtenir càrregues atòmiques, i els orbitals frontera, que són utilitzats en estudis de relació estructural-activitat de fàrmacs.

Les tècniques quàntiques permeten obtenir una visió molt acurada dels sistemes moleculars, i permeten estudiar qualsevol tipus de sistema (inclosos els sistemes reactius com els enzims). No obstant això, tenen dos inconvenients que limiten llur aplicació al camp de les macromolècules: *a)* són molt costoses computacionalment; i *b)* en el seu actual formalisme no proporcionen una visió dinàmica dels sistemes moleculars.

TÈCNIQUES CLÀSSIQUES

Aquestes tècniques tracten els àtoms com si fossin partícules que interaccionen d'acord

amb les lleis clàssiques. Donat que no es consideren explícitament els graus de llibertat electrònics, no poden aplicar-se per a l'estudi de l'enllaç químic, que haurà de ser descrit mitjançant la introducció de diversos paràmetres empírics. D'altra banda, són tècniques molt menys costoses computacionalment, fet que permet la seva aplicació a l'estudi dinàmic de grans sistemes moleculars, inabordables amb les tècniques quàntiques.

Camp de força

L'element comú és la descripció clàssica de les interaccions entre partícules clàssiques fent servir un camp de força, és a dir, un conjunt d'equacions que descriu les diferents contribucions energètiques a la interacció entre partícules del sistema. La utilització d'un camp de força es pot racionalitzar amb la separació de Born-Oppenheimer, ja que els nuclis es mouen dintre del camp mitjà creat pels electrons. Atès que no hi ha una expressió exacta per a aquest tipus d'hamiltonià, una limitació essencial dels mètodes clàssics consisteix en la precisió de la funció empírica emprada per avaluar l'energia potencial, que depèn de com s'hagin definit les diverses contribucions energètiques i de com s'hagi dut a terme la parametrització d'aquests termes energètics. Al seu torn, aquestes característiques depenen de l'aplicació desitjada pel camp de força.

Un camp de força es caracteritza per un conjunt d'atributs, que es detallen breument a la figura 3:

a) Sistema de coordenades. L'energia potencial pot avaluar-se tenint en compte diversos conjunts de coordenades, com ara els modes normals de vibració o bé coordenades internes. Actualment s'acostuma a emprar un conjunt de coordenades internes altament redundant, format per *stret-*

CARACTERÍSTIQUES D'UN CAMP DE FORÇA

Sistema de coordenades

Coordenades normals de vibració, internes
Restricció de coordenades internes (longituds i angles d'enllaç)

Nivell d'expansió

Quadràtic, cúbic, quàrtic, etc.
Termes creuats

Formes funcionals

Longituds: harmònic, Morse, addició de termes cúbics, etc.
Angles: harmònic, termes cúbics, Urey-Bradley, etc.
Torsions: trigonomètric, *out-of-plane*
Van der Waals: Lennard-Jones 6-12, $\exp[-6]$, utilització d'altres potències, etc.
Electrostàtic: coulòmbic, moments dipolars, constants dielèctrica
Termes específics: pont d'hidrogen, polarització

Tipus d'àtoms

Nombre i identificació
Models *united*, *all-atom*

Parametrització i camp d'aplicació

Dades de referència: geometries, vibracions, calors de formació, dades *ab initio*, etc.
Compostos objecte d'aplicació, propietats d'interès, precisió

FIGURA 3. Atributs principals d'un camp de força

ching i *bending* d'enllaços, torsions, i interaccions no enllaçants. A vegades, determinades coordenades poden adoptar valors fixos, com ara les longituds i els angles d'enllaços.

b) **Nivell d'expansió.** Un cop escollit el conjunt de coordenades $\{\zeta\}$, l'energia en un punt determinat (λ''), caracteritzat per un conjunt de valors $\{\zeta\}$, pot expressar-se com una expansió en sèrie de Taylor al voltant d'un mínim local (λ'), (equació 12).

El segon terme, de fet, s'anul·la, per condició de mínim a λ' . Tanmateix, en el segon sumatori, que descriu la contribució qua-

dràtica a l'energia, hi intervenen dos termes diferents: els diagonals, en els quals la diferencial ve donada respecte a una mateixa coordenada (ζ_i), i els termes creuats, que involucren la diferenciació respecte a dues coordenades diferents (ζ_i, ζ_j). Els termes creuats descriuen l'acoblament entre les deformacions en longituds i els angles d'enllaç, o bé l'acoblament entre longituds i torsions, entre angles i torsions, entre longituds i longituds, entre angles i angles i finalment entre torsions i torsions. Tots ells plegats descriuen l'expansió quadràtica general de l'energia tenint en compte les coor-

$$E(\lambda'') = E(\lambda') + \sum \left(\frac{\partial E}{\partial \lambda} \right)_{\lambda'} \Delta \lambda + \frac{1}{2} \sum \left(\frac{\partial^2 E}{\partial \lambda^2} \right)_{\lambda'} (\Delta \lambda^2) + \text{termes superiors} \quad [12]$$

denades internes. Finalment, els termes de tercer ordre i els superiors en l'expansió de Taylor introdueixen els efectes d'inharmonicitat.

Aquests, doncs, són dos dels aspectes que cal considerar en escollir un camp de força. D'una banda, és usual tallar l'expansió en el segon terme, obviant les possibles desviacions del comportament harmònic, fet raonable sempre que es produeixin deformacions petites. D'altra banda, els termes creuats generalment no es consideren en els camps de força dissenyats per a l'estudi de macromolècules, encara que poden ser pre-

sents en camps de força parametritzats per a l'estudi de molècules orgàniques petites.

c) **Forma funcional.** Si bé l'equació 12 limita la forma funcional de les contribucions energètiques pel que fa a potències de $\Delta\lambda$, poden emprar-se altres formes funcionals. Un exemple clàssic és la funció de Morse, en comptes de l'expressió harmònica, per descriure les desviacions en la longitud d'enllaç. De fet, hi ha diverses expressions per als diferents termes que componen el camp de força, tant les interaccions enllaçants (*stretching*, *bending* i torsions) com les no enllaçants (electrostàtic i Van der Waals). La

TERMES ENLLAÇANTS

Stretching

$$K_{AB}(l - l_0)^2$$

l_0 : distància d'equilibri entre els àtoms A i B
 k_{AB} : constant de força de *stretching*

$$D_e \left\{ \exp[-A(l - l_0)] - 1 \right\}^2$$

D_e : profunditat del mínim d'energia

Bending

$$K_{ABC}(\theta - \theta_0)^2$$

θ_0 : angle d'equilibri entre els àtoms A, B i C
 k_{ABC} : constant de força de *bending*

Torsió pròpia

$$\sum_n \frac{V_n}{2} (1 + \cos(n\phi - \alpha))$$

V_n : barrera torsional
 n : periodicitat de la torsió
 α : angle de fase

Torsió impròpia

$$K_{ABCD}H^2$$

H : distància de l'àtom D al pla ABC
 k_{ABCD} : constant de força de torsió impròpia

TERMES NO ENLLAÇANTS

van der Waals

$$4 \pi \epsilon^* \left[\frac{C_{AB}}{R_{AB}^{12}} - \frac{D_{AB}}{R_{AB}^6} \right]$$

ϵ^* : paràmetre dependent de les «dureses» dels àtoms A, B
 C_{AB} , D_{AB} : paràmetres dependents dels radis dels àtoms A, B
 R_{AB} : distància entre àtoms A i B

Electrostàtic

$$\frac{q_A q_B}{\epsilon R_{AB}}$$

q_A : càrrega parcial de l'àtom A
 ϵ : constant dielèctrica

FIGURA 4. Expressions emprades per avaluar les diferents contribucions energètiques en el camp de força

figura 4 mostra de forma esquemàtica algunes de les expressions emprades per avaluar les diferents contribucions energètiques.

d) **Tipus d'àtoms.** La identificació dels àtoms és important en la definició i la parametrització d'un camp de força, ja que la diferenciació de diversos entorns químics s'especifica mitjançant l'assignació de diferents tipus d'àtoms. Pot considerar-se convenient agrupar diversos àtoms com un únic grup (model *united atoms*), com es fa per exemple en la descripció de metilens i metils. Un conjunt petit de tipus d'àtoms proporciona menys complexitat i redueix el nombre de paràmetres, però pot esdevenir insensible als detalls de les característiques químiques.

e) **Parametrització i camp d'aplicació.** La selecció d'un camp de força ve condicionada per les dades de referència que s'han fet servir en l'assignació dels paràmetres per als diversos tipus d'àtoms. Dos camps de força similar parametritzats amb dades diferents poden portar a solucions diverses. Òbviament, la selecció de les dades –experimentals o teòriques– emprades en la parametrització depèn de la previsible aplicació del camp de força. Pot ser fàcil derivar un camp de força acurat per a l'estudi d'un grup de molècules determinat; no obstant això, aquesta tasca esdevé molt més delicada, i el camp de força ha de ser més complex, quan es vol obtenir resultats acurats per a un conjunt divers de molècules.

Mecànica molecular

Un cop seleccionat un camp de força, la seva aplicació a una configuració nuclear determinada permet avaluar l'energia del sistema. Això permet portar a terme una exploració de l'espai conformacional accessible mitjançant algorismes d'optimització de geometria, que constitueix la finalitat més senzilla i directa del camp de força.

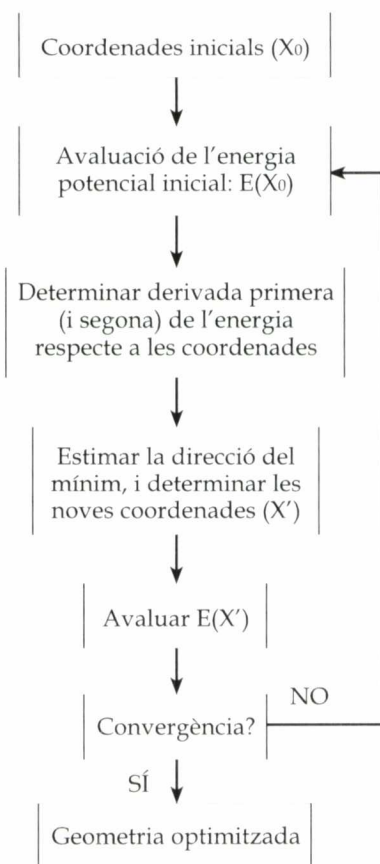


FIGURA 5. Algorisme bàsic dels mètodes de mecànica molecular

Tal com s'indica a la figura 5, donada una configuració nuclear inicial, el camp de força permet obtenir la geometria que minimitza l'energia. L'eficàcia de l'exploració conformacional depèn del procediment emprat en la recerca del mínim d'energia, que pot venir determinat pel gradient d'energia (mètode *steepest descent*), o bé pel gradient corregit amb la informació obtinguda en etapes anteriors del cicle iteratiu de minimització (mètode de gradients conjugats), o bé, finalment, calculant la matriu de derivades segones de l'energia (mètode de New-

ton-Raphson). Tanmateix, una limitació és que l'algorisme de minimització condueixi cap a un mínim local d'energia, sense que s'hagi identificat el mínim global del sistema. Aquest problema s'agreuja molt a mesura que augmenta el nombre de graus de llibertat del sistema, fet que dificulta la possibilitat de dur a terme una exploració exhaustiva de tot l'espai configuracional.

Aquest tipus de metodologia és adient per examinar les conformacions de mínima energia i obtenir la seva geometria, i permet dur a terme el refinament d'estructures moleculars, com també fer una exploració de la flexibilitat conformacional. Una altra aplicació consisteix a analitzar el *fitting* entre biomolècules, com el reconeixement entre lligands i receptors, i facilita l'exclusió d'aquelles interaccions que són estèricament desfavorables o permet estimar les afinitats de diversos substrats per a un receptor comú. Malauradament, aquesta metodologia proporciona únicament una visió estàtica del sistema molecular, que no es correspon amb la situació real, ja que a una temperatura determinada qualsevol sistema pot adoptar un ampli nombre de microestats, el conjunt dels quals defineix les propietats moleculars. A més, cal remarcar el fet que les fluctuacions estructurals degudes a l'energia tèrmica poden desenvolupar un paper primordial pel que fa a la funció de les biomolècules; és necessari considerar aquests efectes per comprendre les implicacions funcionals en determinats processos biològics. Això pot aconseguir-se mitjançant les tècniques de dinàmica molecular i Monte Carlo.

Dinàmica molecular

Aquesta metodologia es basa en l'aplicació directa de les equacions de moviment de Newton o de Langevin per descriure l'evolució temporal d'un sistema. La idea essencial és que el camp de força permet

avaluar no tan sols l'energia per a una configuració nuclear, sinó també la força que actua sobre cada partícula a partir del gradient de l'energia potencial. Fent servir les lleis de la mecànica clàssica, hom pot predir aleshores la posició que ocuparà i la velocitat que tindrà cadascuna de les partícules en un temps determinat. Avaluant successiva-

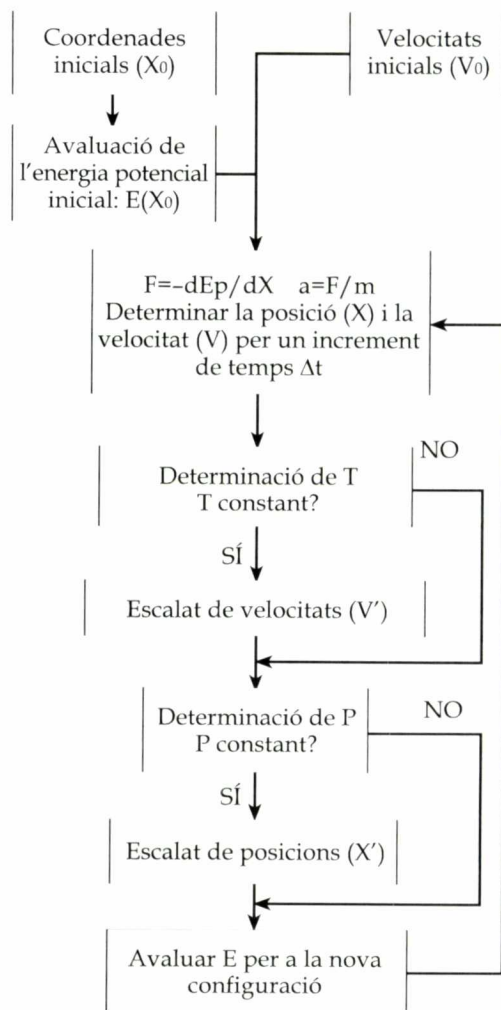


FIGURA 6. Algorisme bàsic dels mètodes de dinàmica molecular

ment en el temps la força sobre cada partícula i estimant la posició i la velocitat futures, és possible obtenir un conjunt d'estructures que representen l'evolució temporal del sistema a una temperatura determinada.

El procediment esquemàtic de la dinàmica molecular es mostra a la figura 6. Per iniciar la dinàmica, cal indicar les coordenades i les velocitats inicials. Les coordenades poden obtenir-se a partir, per exemple, de l'estructura cristal·logràfica determinada per raigs X, mentre que les velocitats es determinen aleatòriament seguint una distribució de Maxwell a la temperatura de referència. Seguidament, es determina l'energia potencial i el seu gradient, la força que actua sobre cada partícula, i per integració de l'acceleració poden estimar-se les noves posicions i velocitats. Posteriorment, pot aplicar-se un factor de correcció a les velocitats i a les posicions per mantenir constant el valor de la temperatura i de la pressió del sistema, segons que la simulació de dinàmica es faci en condicions de temperatura i pressió constant, o d'energia i volum constant.

Dintre de l'esquema anterior, un punt essencial per a l'aplicació pràctica rau en l'elecció dels temps d'integració. Cal pensar que les equacions que cal integrar són complexes; això fa necessari l'ús de mètodes numèrics, per la qual cosa el temps d'integració és molt petit. Tanmateix, aquest temps sempre vindrà determinat per l'escala temporal dels moviments més ràpids en la simulació, que corresponen a l'*stretching* d'enllaços. Per això, el temps d'integració és aproximadament d'un femtosegon. En conseqüència, actualment l'escala de temps real que pot assolir-se en la simulació és com a màxim aproximadament d'uns nanosegons, tenint present que aquesta durada de simulació comporta la realització de milers de càlculs per avaluar l'energia potencial en cadascuna de les diferents configuracions.

A més, ja que les coordenades i les velocitats són assignades arbitràriament, la primera part de la simulació sempre s'ha de considerar com un període transitori, a fi que el sistema pugui equilibrar-se i assolir les condicions desitjades en la simulació. Finalment, per aquest mateix fet, cal tenir present que la dinàmica mai no és unívoca, ja que, dependent de les condicions inicials, el sistema efectuarà una exploració diferent de l'espai configuracional accessible.

Malgrat aquestes consideracions pràctiques, la dinàmica molecular és una tècnica enormement valuosa per determinar un conjunt de configuracions representatives del sistema molecular. Això és possible gràcies al fet que el sistema pot travessar barres energètiques, la qual cosa permet l'estudi de la flexibilitat conformacional de macromolècules amb les salvetats de temps mencionades anteriorment. El coneixement de l'espai configuracional accessible d'un fàrmac permet definir amb precisió els seus patrons de reconeixement farmacofòrics al llarg del temps. Així mateix, permet avaluar el paper de les fluctuacions estructurals en processos d'unió entre biomolècules. A més, pot emprar-se per al refinament de les estructures tridimensionals determinades per difracció de raigs X o bé per ressonància magnètica nuclear. Finalment, pot fer-se servir per estimar els valors mitjans al llarg del temps de diverses propietats del sistema.

Monte Carlo

Aquesta tècnica, a diferència de la dinàmica, no considera el temps com una variable. Bàsicament, el seu fonament rau en la llei de Boltzmann, segons la qual les configuracions energèticament més estables són les que seran ocupades per un nombre més gran de molècules. Així doncs, aquest mètode genera de forma aleatòria una sèrie de

configuracions per a les partícules que componen el sistema molecular (vegeu figura 7). L'energia de cada configuració és avaluada a partir del camp de força. Si l'energia és menor que la corresponent a la de la configuració original, aleshores és acceptada. En cas contrari, encara hi ha certa probabilitat que aquesta nova configuració sigui considerada com a representativa del sistema; aquesta probabilitat serà més gran com menor sigui la diferència energètica de la nova configuració envers la configuració original.

És important remarcar que, malgrat les diferències en el formalisme dels mètodes de dinàmica molecular i de Monte Carlo,

l'exploració de l'espai configuracional de sistemes en equilibri proporcionada per ambdues tècniques és idèntica en el cas que el mostreig sigui suficientment llarg.

Els sistemes d'aplicació i la informació que pot derivar-se dels mètodes de Monte Carlo és molt semblant al que s'ha dit per a les tècniques de dinàmica molecular. El fet que no es consideri el temps com a variable fa que la dinàmica molecular sigui més adient quan l'interès se centra en l'evolució dinàmica d'alguna propietat del sistema. D'altra banda, els mètodes de Monte Carlo permeten establir un control més directe de les variables configuracionals que defineixen el sistema. Així, aquests mètodes han estat molt vàlids per a l'estudi de molècules petites o amb un grau elevat de rigidesa en la seva estructura en fases condensades. No obstant això, en el cas de sistemes grans amb molta flexibilitat, com ara proteïnes, les tècniques de dinàmica molecular són més adients.

Acoblament amb mecànica estadística

L'acoblament de dinàmica molecular i Monte Carlo amb les tècniques de mecànica estadística ha donat lloc al desenvolupament d'uns mètodes computacionals molt sofisticats i de gran impacte per a l'estudi de sistemes biològics. Aquest acoblament permet processar la informació recollida en el mostreig configuracional, a fi d'avaluar les propietats termodinàmiques del sistema. En particular, un punt central d'aplicació ha estat la determinació de diferències d'energia lliure entre dos estats diferents del sistema. Així, si 0 i 1 designen aquests dos estats, la diferència d'energia lliure ve donada per l'equació 13, on k és la constant de Boltzmann, T és la temperatura, E és l'energia del sistema, i β és l'invers del producte kT . Els parèntesis representen el valor mitjà ponderat amb el factor de Boltzmann.

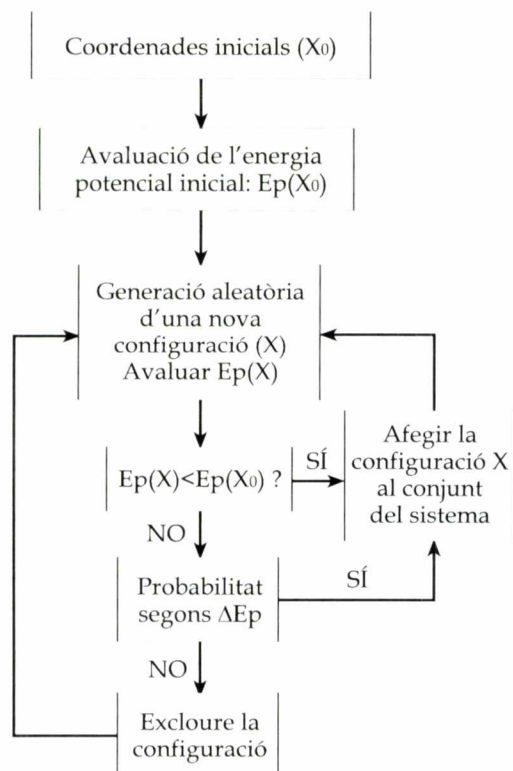


FIGURA 7. Algorisme bàsic dels mètodes de Monte Carlo

$$\Delta A = -k T \ln \frac{\langle \exp[-\beta E_1] \rangle}{\langle \exp[-\beta E_0] \rangle} \quad [13]$$

Des d'un punt de vista pràctic, l'expressió anterior té limitacions, ja que el valor mitjà de l'exponencial presenta una convergència lenta i, a més, l'energia lliure es determina a partir de dos valors molt elevats. S'han proposat diversos mètodes per superar aquests inconvenients, com ara la pertorbació d'energia lliure, el potencial de força mitjana i la integració termodinàmica. Per exemple, en el cas de la pertorbació d'energia lliure es defineix un paràmetre d'acoblament entre els dos estats, que permet transformar l'estat 0 en l'estat 1 de forma gradual. Així, en cada etapa d'aquesta transformació l'hamiltonià del sistema pot definir-se segons el paràmetre d'acoblament, tal com s'expressa a l'equació 14. D'aquesta manera, per a $\lambda=0$ recuperem l'hamiltonià de l'estat 0, mentre $\lambda=1$ correspon a la situació de l'estat 1.

$$H_\lambda = (1 - \lambda) H_0 + \lambda H_1 \quad [14]$$

El fet de dur a terme aquesta transformació modificant gradualment el valor del paràmetre d'acoblament permet avaluar la diferència d'energia lliure com la suma de les variacions per a cadascuna de les etapes graduals. L'expressió corresponent a la pertorbació d'energia lliure (equació 15) indica que per a cada etapa intermèdia cal determinar la diferència d'energia per als estats successius λ i $\lambda+\Delta\lambda$ fent servir el mostreig configuracional corresponent a l'estat λ .

$$\Delta A = \sum_{\lambda=0}^{\lambda=1-\Delta\lambda} R T \ln \langle \exp[-(E_{\lambda+\Delta\lambda} - E_\lambda) / RT] \rangle_\lambda \quad [15]$$

Cal remarcar que la naturalesa del paràmetre d'acoblament i els estats intermedis de la transformació entre estats són irrelevants per al càlcul de ΔA , ja que es tracta d'una funció d'estat. A més, aquestes tècniques

permeten determinar amb una precisió sorprenent la variació d'energia lliure associada a diversos processos bioquímics. De fet, una revisió de la literatura recent mostra amb tota claredat l'enorme impacte d'aquestes tècniques per a l'estudi quantitatiu de sistemes biològics, tals com les diferències d'energia lliure de *binding* entre dos substrats o l'efecte de mutacions sobre l'estabilitat de complexos. No obstant això, els resultats són molt sensibles a les condicions de simulació, i cal examinar amb detall el protocol de simulació per tal d'evitar l'obtenció de resultats erronis.

L'enorme potencial d'aplicació d'aquestes tècniques sorgeix quan es consideren en combinació amb cicles termodinàmics, tals com els que s'indiquen a la figura 8. El diagrama de la part superior representa el cicle emprat per determinar l'energia lliure de solvatació (ΔG_{solv}) d'una molècula. En aquest cas, el cicle consisteix en dues etapes d'anihilació de l'espècie química, és a dir, a efectuar la transformació de la molècula a una entitat nul·la en fase gas i en solució. Atès que l'energia d'interacció entre aquesta entitat imaginària amb les molècules de solvent és zero, l'energia lliure de solvatació pot avaluar-se per la diferència de treballs necessaris per anihilar el compost en les dues fases.

L'exemple mostrat en el diagrama central de la figura 8 permet avaluar la diferència en el procés d'unió de dos lligands (L1 i L2), que poden diferir en llur composició i estructura química, i que interaccionen amb el mateix receptor. En aquest cas, la diferen-

cia entre les energies lliures d'interacció vindria donada per la diferència entre els valors $\Delta G1$ i $\Delta G2$, que són difícils de determinar computacionalment. No obstant això, mitjançant l'ús del cicle termodinàmic,

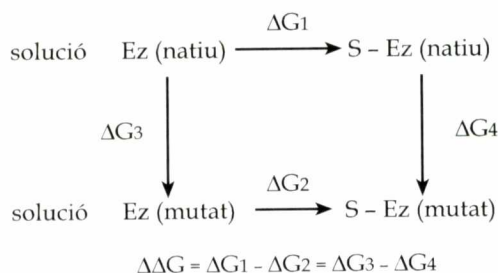
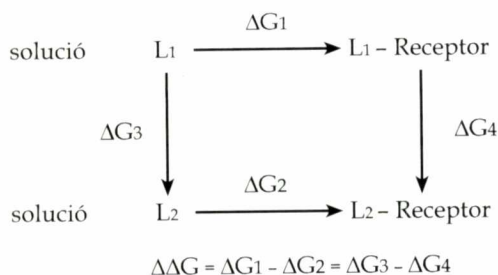
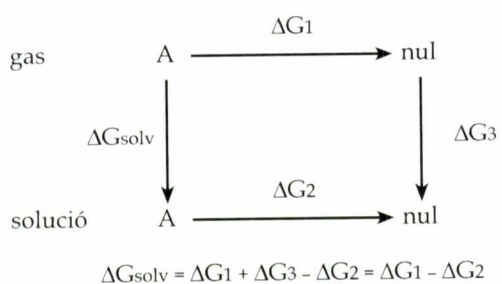


FIGURA 8. Utilització de les tècniques de perturbació d'energia lliure dintre del marc de cicles termodinàmics per avaluar energies lliures de solvatació, diferències d'energia lliure de *binding* o l'efecte de mutacions en enginyeria d'enzims

aquesta diferència pot avaluar-se a partir dels valors ΔG_3 i ΔG_4 , que representen la variació d'energia lliure en el procés reversible de transformar un lligand en l'altre, bé en solució o bé units al receptor.

L'exemple mostrat a la part inferior és semblant a l'anterior, però en aquest cas correspondria a l'efecte diferencial que té la

introducció de mutacions puntuals en residus de l'enzim sobre la seva interacció amb un substrat. De nou, la diferència en la capacitat per unir-se al substrat pot estimar-se a partir dels treballs reversibles necessaris per transformar l'enzim natiu en la forma variant, tant per a l'enzim en solució com per a l'enzim un cop s'ha unit al substrat. Sens dubte, aquest tipus de procés té moltes implicacions sobre l'enginyeria de proteïnes i el disseny d'enzims amb característiques específiques, com ara modular l'especificitat per substrats.

Aplicacions i limitacions

Les tècniques clàssiques són molt potents per tal de descriure grans sistemes moleculars. Són, per tant, molt emprades en l'estudi de les interaccions entre macromolècules, o entre macromolècules i petits lligands. També són els mètodes que cal emprar per tal d'estudiar processos de canvis conformacionals com ara el plegament/desplegament de proteïnes, o el *fitting* entre DNA i proteïnes.

Aquests mètodes presenten tres limitacions bàsiques, que estan relacionades amb: a) la impossibilitat de les tècniques clàssiques per estudiar reaccions químiques; b) les aproximacions fetes en la funció d'energia potencial; i c) l'extensió de les simulacions. La primera fa impossible l'ús de les tècniques teòriques en l'estudi de la reactivitat enzimàtica, o en l'estudi de les modificacions químiques de les proteïnes; la segona introdueix errors sistemàtics en els càlculs; mentre que la tercera dóna lloc a errors estadístics. Nombrosos grups estan treballant actualment en l'elaboració de camps de força més acurats i en la derivació de paràmetres més refinats. En aquest punt, hi ha cada vegada més interès a fer la parametrització no utilitzant dades derivades d'experiments en fase gas o de càlculs *ab ini-*

ció per molècules aïllades, sinó fent servir dades en solució o bé dades cristal·logràfiques. Tanmateix, la descripció més acurada dels termes no enllaçants, com ara la interacció per pont d'hidrogen o el tractament dels efectes de polarització, és cada vegada més important, com també ho és la inclusió dels efectes electrostàtics de llarga distància.

Respecte al problema de l'extensió, simulacions de l'ordre d'1ns són relativament assequibles amb els actuals mitjans computacionals. No obstant això, és obvi que el marge de processos biològics que pot ser considerat amb aquesta escala de temps és relativament reduït. Així, caldrien simulacions més extenses per caracteritzar aquelles fluctuacions dinàmiques que comporten canvis conformacionals de llarga escala, detalls de cinètica enzimàtica o bé d'unió de substrats, i especialment el problema del plegament de proteïnes, per al qual s'han proposat diverses estratègies fent servir models més o menys simplificats. Respecte a la limitació intrínseca de les tècniques clàssiques en l'estudi de sistemes reactius, la solució és integrar els mètodes quàntics i clàssics en les anomenades tècniques mixtes.

TÈCNiques MIXTES

Aquest grup de tècniques constitueix un camp de treball molt actiu en el moment actual, que sens dubte tindrà, cada cop més, un major impacte en l'estudi de les propietats estructurals i funcionals de les proteïnes. Tal com es va comentar anteriorment, el formalisme bàsic consisteix a distribuir el sistema molecular en dos subsistemes: un d'ells es descriu amb tècniques quàntiques, i l'altre és tractat clàssicament (vegeu figura 9). La descripció de la part clàssica del sistema pot fer-se utilitzant partícules discretes, caracteritzades per les seves càrregues parcials (q) i els paràmetres de duresa (ϵ^*) i radi (σ) de van der Waals, o bé un continu polaritzable, caracteritzat per la seva constant dielèctrica (ϵ). Aquesta separació es fa de manera que el subsistema quàntic descriu la part del model material on té lloc l'objecte d'interès, mentre que la resta de l'entorn és descrita clàssicament. En el cas de l'estudi de l'activitat catalítica d'un enzim, per exemple, el subsistema quàntic estaria format pel substrat i aquells residus que intervinguen directament en el procés reactiu, mentre que la resta de la proteïna amb les molècules de solvent constituïria el subsistema clàssic.

En el cas dels mètodes mixtos quàntics-discrets, el subsistema quàntic es descriu actualment fent servir hamiltonians semiempírics o bé mètodes de densitat funcional, encara que s'han publicat alguns resultats obtinguts per Hartree-Fock. L'energia total del sistema pot expressar-se com la suma de les energies associades al subsistema quàntic, clàssic, i al terme d'interacció entre ambdós (equació 16). El terme d'interacció s'avalua normalment com l'addició de dues contribucions, que corresponen a l'energia electrostàtica entre nuclis i electrons del subsistema quàntic amb les càrregues parcials emprades per a la distribució

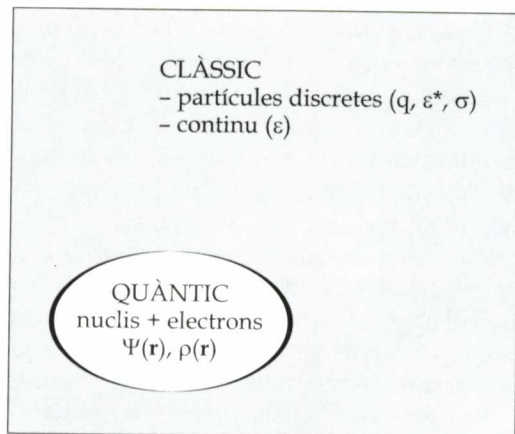


FIGURA 9. Partició del sistema en un subsistema quàntic i un subsistema clàssic

de càrrega de les partícules clàssiques, i l'energia d'interacció de Van der Waals entre els àtoms i les partícules clàssiques (equació 17). En l'equació 17, N és el nombre d'orbitals moleculars doblement ocupats, M és el nombre de nuclis, i S és el nombre de partícules clàssiques.

$$\hat{H} = \hat{H}_{QM} + \hat{H}_{MM} + \hat{H}_{QM-MM} \quad [16]$$

$$\hat{H}_{QM-MM} = \sum_{s=1}^S \sum_{i=1}^{2N} \frac{e q_s}{R_{si}} + \sum_{s=1}^S \sum_{m=1}^M \frac{Z_m q_s}{R_{sm}} + E_{vw} \quad [17]$$

L'energia, doncs, pot obtenir-se a partir de la corresponent equació de Schrödinger, fent servir l'hamiltonià efectiu descrit a l'equació 16. Cal tenir present en aquest punt que la representació real del sistema ve donada pel conjunt de configuracions assequibles a una temperatura determinada, i que aquest mostreig pot obtenir-se fent servir els mètodes de dinàmica molecular o de Monte Carlo. Per a cada configuració de les partícules serà necessari, doncs, dur a terme el càlcul de la funció d'ona del sistema i de la seva corresponent energia. Això comporta un enorme cost computacional, fet que justifica la utilització de mètodes semiempírics o de funcional densitat en la descripció de la part quàntica.

Pel que fa als mètodes mixtos quàntics-continus, la representació discreta de l'entorn se substitueix per un continu dielèctric, que es polaritza davant el camp elèctric generat per la distribució de càrrega del subsistema quàntic. El camp de reacció generat per l'entorn polaritza al seu torn la densitat de càrrega del subsistema quàntic, i aquest procés continua fins arribar a l'autocoherència. L'energia total del sistema, aleshores, pot determinar-se a partir d'un hamiltonià efectiu (equació 18), on \mathbf{R} simbolitza el camp de reacció. Cal notar que, com que \mathbf{R} depèn de la funció d'ona del subsistema

quàntic i aquesta ve determinada per la magnitud de \mathbf{R} , és necessari dur a terme un procés iteratiu. La forma explícita que adopta l'operador \mathbf{R} depèn del model matemàtic escollit. En alguns mètodes es defineix fent servir una expansió multipolar, mentre que altres formalismes descriuen el camp de reacció a partir d'una distribució de càrrega sobre la superfície de separació dels sistemes quàntics i clàssics.

$$(\hat{H}^0 + \hat{\mathbf{R}}) \psi = E \psi \quad [18]$$

L'eliminació dels graus de llibertat de les molècules que formen el sistema clàssic permet assolir nivells de teoria més elevats en la descripció del subsistema quàntic. No obstant això, cal tenir present que l'efecte de les interaccions específiques amb les molècules clàssiques es perden completament. Si bé aquesta aproximació no sembla dramàtica pel que fa a l'estudi de molècules en solució, on aquests mètodes prediuen energies lliures de solvatació molt acurades, no són adequats per a l'estudi de processos en medis amb una anisotropia molt elevada, com és el cas dels processos químics que tenen lloc en l'interior de les proteïnes.

ALTRES TÈCNIQUES PER A L'ESTUDI DE BIOMACROMOLÈCULES

És obvi que el ventall de tècniques disponibles per a l'estudi de macromolècules biològiques comprèn un conjunt de mètodes més variat que els que s'han descrit fins ara. No és possible fer una descripció de totes les tècniques, però sí que considerem necessari esmentar-ne almenys algunes, a fi de proporcionar una visió més completa.

Dintre d'aquest context, voldríem fer referència als programes desenvolupats amb el propòsit específic d'abordar l'estudi del *docking*, és a dir, la predicció de les estructures dels complexos intermoleculars formades per la unió de substrats/l·ligands a proteïnes. Aquests programes generen un nombre molt elevat de possibles estructures d'interacció, i efectuen una valoració de cada estructura a fi de determinar quines són veritablement significatives per comprendre el procés d'unió. El problema més important en aquests estudis és, sens dubte, l'elevat nombre de graus de llibertat, que ve determinat per la flexibilitat conformacional de les espècies interaccionants. De fet, només es consideren els graus de llibertat del lligand, mentre que l'estructura de l'enzim o el receptor se suposa que és completament rígida.

Un altre conjunt de mètodes molt emprats són aquells basats en la resolució de l'equació de Poisson-Boltzmann per diferències finites, a fi d'examinar les propietats electrostàtiques de les macromolècules en solució. En aquests mètodes, es fa una descripció des de la configuració atòmica de la macromolècula, representada com un conjunt de partícules clàssiques caracteritzades per càrregues parcials i paràmetres de Van der Waals. La macromolècula està separada del medi aquós per una superfície definida segons els radis atòmics, i, per tant, està adaptada a la seva forma tridimensional. Normalment, es defineix una constant dielèctrica de 2 a l'interior de la macromolècula, a fi de considerar els efectes de polarització electrònica, mentre que en el medi exterior s'assigna una constant dielèctrica de 80, que fa possible la consideració dels efec-

tes de força iònica, descrits d'acord amb la teoria de Debye-Hückel. Aquest procediment permet estimar l'energia lliure electrostàtica de la molècula en solució. Les aplicacions són diverses i comprenen aspectes com ara: determinar propietats àcid/base per als residus en proteïnes, analitzar el procés d'unió de biomolècules, avaluar l'energia lliure de solvatació, o bé examinar l'efecte de mutacions sobre l'estabilitat de macromolècules.

Finalment, cal fer menció d'un conjunt de tècniques anomenades «estadístiques» que són emprades per tal de predir el comportament i l'estructura de proteïnes. Aquestes tècniques funcionen en tres etapes: a) anàlisi de l'estructura i les propietats d'un gran nombre de proteïnes conegudes; b) formulació de regles empíriques entre la seqüència i les característiques de les proteïnes conegudes; i c) aplicació de les regles per tal de predir l'estructura o les propietats de proteïnes de seqüència coneguda. Mètodes com els de Chou-Fasman, o els de predicció de perfils hidròfils, són exemples d'aquests mètodes.

BIBLIOGRAFIA

- SZABO, A.; OSTLUND, N. S. (1982). *Modern quantum chemistry*. McMillan, Nova York.
- LEACH, A. R. (1996). *Molecular modelling. Principles and applications*. Longman, Essex.
- HEHRE, W. J.; RADOM, L.; SCHLEYER, P. v. R.; POPE, J. A. (1986). *Ab initio molecular orbital theory*. John Wiley and Sons, Nova York.
- ALLEN, M. P.; TILDESLEY, D. J. (1989). *Computer simulations of liquids*. Clarendon Press, Oxford.
- MCCAMMON, J. A.; HARVEY, S. C. (1987). *Dynamics of proteins and nucleic acids*. Cambridge University Press, Cambridge.